

# White Paper

Report ID: 109698

Application Number: HG-50047-13

Project Director: Brett Barney

Institution: University of Nebraska, Board of Regents

Reporting Period: 1/1/2014-12/31/2015

Report Due: 3/31/2016

Date Submitted: 4/1/2016

## White Paper

### Introduction

The editor who wishes to use the current TEI *Guidelines* (P5: *Guidelines for Electronic Text Encoding and Interchange*: Version 3.0.0) to create a genetic edition faces several daunting challenges and is advised to proceed deliberately and to limit her or his expectations based on a careful assessment of the current provisions for genetic editing offered by the *Guidelines* and the feasibility of leveraging the markup in specific ways.

“Genetic editing” is not a monolithic practice, and significant differences in theory, emphasis, and procedure exist among the major editorial schools/approaches usually grouped under this banner. For many editors, however, a diplomatic transcription is a key component of a genetic edition. The provisions for doing such a transcription constitute the bulk of changes introduced in the major revision of chapter 11, “Representation of Primary Sources” and are currently better developed than the provisions for indicating change over time. Furthermore, the *Guidelines*, as they are currently written, subtly steer editors—even those who may have no overt devotion to genetic editing *per se*, or even to diplomatic editing itself—toward diplomatic transcriptions (via `<sourceDesc>`) instead of textual ones (via `<text>`).

The methods for creating a textual encoding within the `<text>` element are highly developed, the result of over two decades of the TEI’s evolution. For this reason, we recommend that unless you have well-articulated reasons for needing a diplomatic transcription of your material you use the well-developed provisions for encoding textual meaning inside the `<text>` element and forego using `<sourceDoc>` altogether.

### Considerations for Those Doing Diplomatic Transcriptions

If a diplomatic transcription is central to your purposes, however, you should, at the outset of your project, devote thoughtful deliberation to establishing several basic aspects of your encoding protocol that will have important consequences for later reuse, processing, etc. We emphasize the importance of this planning step especially because the *Guidelines* provide very little guidance for conducting such deliberations, and their silence on such matters might lead one to underestimate the importance of basic decisions or to overlook their implications.

One of the most fundamental choices you make will be whether to do an “embedded” or “parallel” transcription (or to do some combination of the two). Unfortunately, our experiences on the Diachronic Markup Project have not led to any clear conclusions that would allow us to recommend any particular approach over the others without regard to project-specific circumstances. However, the following cautions may help as you deliberate among them:

- Many of the elements provided for diplomatic transcription (e.g., `<line>` and `<zone>`) cannot be used within `<text>`. And while most of the elements provided for textual transcription (e.g., `<add>` and `<choice>`) can be used within `<sourceDoc>`, `<subst>` cannot.

- If these restrictions lead you to use both <sourceDoc> and <text>, you will need to decide which elements to use in each and whether (and how) you will cross-reference the two.
- Avoid nesting <zone> within <zone> and <line> within <line>, as such structures are likely to be difficult to process.
- If you have as a goal to map parts of the transcription to an image, provide coordinates for each block marked in the diplomatic transcription.

### Considerations for Those Wishing to Encode Temporality

The *Guidelines*' principle instrumentality for encoding diachronic information is via <listChange> (in combination with <change> elements as children, cross-referenced via @change attributes on the various elements used to mark up blocks of text in the transcription proper). This method was developed for the encoding of "revision campaigns" (i.e., the inferred rounds/sessions during which the portions of a single text were written), and the discussion of <listChange> in the *Guidelines* assumes this as the primary scenario for the use of these elements/attributes. Although it is possible to use this mechanism to encode other "levels" of time-dependent relationships (i.e., either micro-scale or macro-scale diachronies), we do not recommend doing so. As the *Guidelines* offer no clear direction for encoding either fine-grained intra-document diachronies or inter-document diachronies, we advise that one defer undertaking such a task until such time as the TEI Council offers more direction. If, however, one has strongly compelling reasons to undertake an encoding of a detailed sequence of inscriptions within an individual document, we recommend using a carefully devised set of custom elements, added to your project's schema via the extension mechanisms provided by the Roma tool and ODD. For inter-document diachronies, two possible basic frameworks are worth considering: RDF and "directed acyclic graphs," as outlined in "Section 4: The dossier level" of "An Encoding Model for Genetic Editions," the draft recommendations written by the TEI's Workgroup on Genetic Editions ([http://users.ox.ac.uk/~lou/wip/geneticTEI.doc.html#index.xml-body.1\\_div.4\\_div.1](http://users.ox.ac.uk/~lou/wip/geneticTEI.doc.html#index.xml-body.1_div.4_div.1)). Either approach will require further refinement on the part of the implementer to specify such things as types of relationships, certainty, and alternative theories. No already existing RDF vocabulary appropriate to describe genetic textual relationships seems to exist; likewise, nothing like a universally applicable set of values for the components of a directed graph have been devised. Thoughtful development of such, undertaken with input from a variety of digital editing projects, would be most valuable, however.